心理科学进展 2018, Vol. 26, No. 9, 1617–1623
Advances in Psychological Science DOI: 10.3724/SP.J.1042.2018.01617

场景知觉过程中的动作意图识别*

康廷虎 薛 西
(西北师范大学心理学院视觉认知实验室, 兰州 730070)

摘 要 场景即我们生活于其中的真实环境，社会场景是其重要组成部分。在社会场景知觉的研究中，动作意图的识别既受场景背景信息的影响，也与动作的客观对象有关。因此，研究者可以根据背景−刺激物、刺激物−刺激物关系，探索动作识别的影响机制；另一方面，也可以根据场景的语义约束和物理限制，依据合理动作原则及其伴随的生理指标检测并识别动作意图。在机器视觉研究领域，计算机识别模型为社会场景中动作意图的检测和识别提供了新的视角。在未来的研究中，研究者需要考虑真实场景中动作意图识别能力的发展、动作意图识别的个体差异和文化差异等问题。

关键词 社会场景；动作意图；场景知觉；计算机识别模型
分类号 B842

1 引言

动作意图既可以指动作的立即结果，也可以指导致动作的高级动机(Catmur, 2015)。动作意图的理解对我们的生活至关重要。对婴儿而言，动作不仅是其获得感性认识的手段，也是与他人进行社会互动的主要方式(陈亚萍，李晓东，2013)；对成人而言，正确理解动作意图是个体在社会生活中与他人进行有效交流的基础(den Ouden, Frith, Frith, & Blakemore, 2005; Satpute et al., 2005)。因此，动作意图识别已经成为计算机科学和心理学领域内的主要研究问题之一(Catmur, 2015; Yao & Fei-Fei, 2010)。

人们对动作意图的理解，不仅依赖于动作本身；也有赖于动作所发生的真实生活场景。场景(scene)是真实世界中各个分散的刺激物及其背景构成的、具有语义一致性的视觉图景(Henderson & Hollingworth, 1999; 白学军，康廷虎，闫国利，2008)。场景知觉关注人如何知觉和加工复杂的真实环境信息(王福兴，田宏杰，申继亮，2009)。有研究者认为，真实世界中的场景知觉包括对视觉输入的感觉过程和认知过程，比如对场景梗概、空间布局与规模等信息的快速获取，以及场景中的距离知觉、有意义对象的视觉搜索、场景的表征及注意分配等(Henderson, 2005)。社会场景(Social scene)，即有人存在的场景，是场景的主要类型之一(Cerf, Harel, Einhäuser, & Koch, 2007)。在社会场景知觉的研究中，对人与人之间互动关系的探索，特别是对人与人之间行为互动背后隐含的心理语义，即动作意图的探索是其重要内容。基于场景知觉的动作意图研究，需要注意场景的背景(background)及其包含的物体(objects)两个部分。场景中的背景是指宽广的、静止的表面和结构；场景中的物体则是指比例较小的不连续物体(王福兴等, 2009)。在真实场景中，场景的背景与其包含的刺激物，以及分散的刺激物之间存在着某种依赖或共现关系，从而构成了刺激物-刺激物关系、刺激物-背景关系，而其都对动作意图识别有重要影响(Bonchek-Dokow & Kaminka, 2014; Yao & Fei-Fei, 2010)。

在日常生活中，对于视觉正常者而言，80% ~ 90%的外界信息来源于视觉通道(康廷虎，白学军，2008)；而且，有许多研究者已经从视觉加工的角度研究动作意图(Bonchek-Dokow & Kaminka, 2014; Sartori, Bechio, & Castiello, 2011)。本文主要对视

收稿日期：2017-05-02
* 国家社会科学基金青年项目(13CSH074)的支持和甘肃省体育卫生与健康教育美育国防教育专项任务项目(项目编号：77)支持
通信作者：康廷虎, E-mail: kangyan313@126.com

觉信息加工的研究成果进行梳理与分析，并基于社会场景知觉的研究，综述动作意图检测、意图分类和意图推论等动作意图识别的相关研究进展(Park, Lee, Lee, Chang, & Kwak, 2016)。

2 场景中的动作意图研究

在社会场景知觉的研究中，动作意图是其中的重要内容。对婴儿而言，各种运动、动作的发展是其活动发展的直接前提，也是其心理发展的外在表现(李红，何磊，2003)。动作不仅是婴儿获得感性认识的手段，也是其与他人进行社会互动的主要方式，尤其是对于前语言阶段的婴儿，动作理解可以看作是一种前心理理论，对促进婴儿其他社会认知能力的发展具有重要的意义(陈亚萍，李晓东, 2013)。因此，理解动作意图对个体心理发展以及人际交往与沟通都具有重要意义(Cacippo, Berntson, & Decety, 2010)。尽管人们所看到的动作流是极其复杂的，但是从婴儿期开始，个体就可以轻松地处理意图相关的动作。人们自发地根据意图边界对动作进行分段，得到关于行为表现者意向性的系统判断，并利用对行为表现者特定意图内涵的判断指导自己的观察、推论和后续动作(Baldwin & Baird, 2001)，这表明人们从很小的时候就可以对行为意图进行识别。另外，从进化的角度来看，在危险场景中准确识别对自己具有威胁的行为对其生存及发展具有重要的适应性作用。因此，对动作隐含的意图进行研究就显得尤为重要。

Catmur (2015)认为，动作意图既可以指动作的立即结果，也可以指导致动作的高级动机。对动作意图的识别可以帮助人们预期他人行为的结果，也可以帮助人们理解动作发出者的意愿和目标。Sukthankar, Geib, Bui, Pynadath 和 Goldman (2014)认为动作意图识别是一种认知他人计划、目的的能力，使得人类可以推论行为表现者正在做什么、为何这样做以及接下来会怎么做。主体可以凭借对动作意图的识别，获得对他人目标的理解，并可以预测其后期动作及运动轨迹(Bonchek-Dokow & Kaminka, 2014)。需要强调的是，研究者往往关注的并不是所有的动作，而仅仅是可以作为意图识别中介的动作，即意图性动作。Bonchek-Dokow 和 Kaminka (2014)认为“意图性动作”是指可能带来某种期望的最终状态的动作，在这一过程中动作作为旨在实现某种隐含意图的中介存在。在这一概念中有三个关键词：动作(action)、意图(purpose)、最终状态(final state)。这三个关键词将意图性动作与其他术语进行区分。其中，“动作”表示“意图性动作”导致了客观世界的某种变化，而识别意图时又可以将“动作”作为中介；“最终状态”指的是动作序列导致了怎样的最终结果状态；“意图”这一术语与期望的最终状态相关。

在实际识别动作意图的过程中，往往需要使用可直接得到的各类信息，以此推论行为者动作的隐含意图，进而帮助人们识别动作的意图。基于这一思路，研究者试图利用得到的生物信息进行推论。Choi (2013)设计了情境意识系统(situational awareness system)用以检测图像中异常行为的意图。除此之外，还有研究认为“功能可见性”在动作意图预测中扮演着重要角色(Bonchek-Dokow & Kaminka, 2014)。这一概念首先由 Gibson (1977)引入，并认为一个对象的属性和它提供的功能相对应，一个物体或环境会暗示其物理属性的所有可能性。如，办公室的座椅表明其可以用来坐着休息；围巾的保温属性说明其可以用来保暖，其厚重属性暗示其可以被折叠以用来靠枕。每个动作序列都有其引起的状态结果，每一个提取的状态结果也有诱导其产生的动作序列，这使人们在谈及功能可见性时，就能够预期或利用可能的目标状态。也就是说，当试图识别隐含于动作序列之中的意图时，人们可以从可能的目标状态出发，利用环境中可得到的功能可见性而实现其目的。

也有研究试图对动作意图进行分类。比如，基于真实场景的特点，合作情境中理解的同伴意图对于将其动作与共同目标匹配是必不可少的(Sebanz, Bekkering, & Knoblich, 2006)；而理解在冲突情境中对手意图对于免遭他人行为对自己的伤害也是同样重要的(Ruys & Aarts, 2010)。与之不同的是，另有研究者从意图本身出发，将社会意图分为合作意图和竞争意图。合作意图是与同伴合作共同完成某个任务，而竞争意图的目标则是与对手竞争以率先完成某个任务(Sartori et al., 2011)。前者指向合作行为，而后者指向竞争行为。在动作的具体表现特征方面，竞争意图可能由于其竞技性质而导致其所引导的动作在速度上与合作意图所引导的动作有所区别。

3　语义关系与动作意图研究

动作意图研究与语义的获得具有密切的联系。"语义(semantics)"来源于盎格鲁-撒克逊语，迄今为止仍与德语动词"meinen"相关，而这个词汇指的是思考(think)或意向(intend)，在这个意义上动作意图是与语义有关的。语义是指消息发出者与接受者对信息意义的理解，以及通过具体背景线索做出的推论(Ziaeefard & Bergevin, 2015)，其实质是对客观刺激对象及其相互关系的理解(Muehlhaus et al., 2014)，而动作意图研究中强调对象不可独立存在，在这一点上，语义与动作意图是一致的。根据 Henderson (2005)对场景的界定，真实场景实际上是包含了背景和具体对象的。与之相似的是，对动作意图的识别，要依赖于动作的发起者(人)，以及动作的对象(可能是人，也可能是非人的其他客观刺激)。因此，从背景和对象的角度考虑，真实场景中同时包含背景和物体，两者并不是孤立存在的，可能会表现出背景–刺激物关系、刺激物–刺激物关系，这两种关系对动作意图的觉察与识别同样具有重要意义(Delaitre, Sivic, & Laptev, 2011)。

3.1　背景–刺激物关系对动作意图识别的影响

在真实的场景中，背景和刺激物之间可能存在共现关系，比如，停车场作为背景，往往是与停放的车辆存在于同一个时空之中。那么，无论是对场景的识别，还是对场景中刺激物的识别，都可能会受到这种共现关系的影响。在包含人的动作的社会场景中，同样，也会因为人与场景背景之间的共现关系，而使场景中人的动作的识别，或者场景的识别受到背景–刺激物关系的影响。

Friedman (1979)指出，人们对场景诊断刺激的优先识别，反过来会促进场景识别。比如，人们对菜刀的优先识别，可能会促进对"厨房"场景的识别。那么，在社会场景中，如果观察者能够对场景中某个人的动作做出识别，是否会影响对动作对象以及整个场景的识别呢？比如，当看到某个人的投篮动作，我们可能会更容易判断这是在篮球场，或者预测防守队员的位置及其动作。因此，对动作意图的研究，需要考虑场景与刺激对象的匹配情况。而且，有些动作可能只在特定的场景中出现(Ziaeefard & Bergevin, 2015)，其同样有可能成为识别某一种社会场景的诊断刺激。一般而言，游泳只在游泳池进行，即使在其他环境中察觉到了游泳的动作特征，人们也不会得到"场景中的人在游泳"的识别结果。另一方面，对场景的识别也可以促进对包含在其内的刺激物的识别。Henderson (2005)的研究发现，场景背景对于刺激信息的视觉搜索及注意分配具有引导性作用。具体而言，场景中作为背景的各项线索会影响人们对动作意图的理解(Ziaeefard & Bergevin, 2015)。如，汽车出现在停车场场景中和出现在公路场景中会诱导人们对驾驶者行为意图的不同预期。如果汽车出现在停车场内，人们可能会倾向于认为驾驶者是想停车；而在公路场景下，人们更可能认为驾驶者是在进行行驶操作并且其行为指向某一目的地。

此外，人们有关场景上下文背景的序列性知识，对于动作意图分析而言也是极其重要的。它可以作为语义信息帮助人们预测动作及其结果(Oliva & Torralba, 2007)，即帮助人们推断动作意图。如，"将某物从箱中取出"和"将某物放入箱中"这两个动作序列具有不同的隐含意图，但是两者都具有"手抓住某物"这一动作，此时，在区分并识别这两种不同的隐含意图时，对具体动作序列的理解就显得尤为重要。

3.2　刺激物–刺激物关系对动作意图识别的影响

在场景中识别动作的隐含意图时，人的具体动作可能是与操作对象相联系的，不同的动作与其动作目标对象之间是相互联结的。因此，识别动作的隐含意图时，对该动作涉及的关联对象的识别可以帮助人们理解动作意图。比如，在行为者动作特征不变的情况下，如果该动作特征出现在草坪场景中，同时伴随行为者出现的对象为足球，人们倾向于推断行为者的动作特征是为了踢足球做出的；然而，如果该动作发生在羽毛球场内，伴随出现的对象为羽毛球球网，人们可能做出"行为者的动作意图是打羽毛球"的推论。

此外，人的身体姿势和行为对象可以作为彼此交互影响的刺激信息(Desai, Ramanan, & Fowlkes, 2010; Delaitre et al., 2011)。也就是说，对于其中一个刺激物的识别可以促进对于另外一个刺激物的识别。比如，在板球运动中，如果没有察觉到板球，很难对行为表现者正在使用板球棒进行防御性击球的动作进行精准判断；同样的，如果没有识别到行为表现者的击球动作，也很难注意到在空间

尺寸上相对较小的板球。在计算机视觉的相关研究中，有研究者基于刺激物–刺激物之间的相互关系对于彼此识别的易化，提出计算机模型以解释场景中物体的识别(Yao & Fei-Fei, 2010)。

4 动作意图识别及相关计算机模型构建

在动作意图的研究中，意图之类的心理学概念一般都是很难直接测定的，因此需要通过对可观测的其他指标的测量与识别，从而实现对动作意图的识别。不同研究者采用不同指标作为中介对动作意图进行研究。

4.1 动作意图识别的指标及原则

许多研究者采用动作发生时伴随的生物指标作为中介。如, Carpenter, Akhtar 和 Tomasello (1998)在对婴儿模仿动作的隐含意图进行研究时，使用情感声音和面部证据作为中介; Jang, Lee, Mallipeddi, Kwak 和 Lee (2013)在特定行为情景中，对行为者基于某种任务的动作和无特定任务的动作进行研究时，使用注视点数目、注视时长、瞳孔大小变化、瞳孔大小变化梯度、眨眼变化等眼动指标试图考察并研究行为者的动作意图。

此外，合理动作原则(Principle of Rational Action)也是动作意图研究的重要理论基础(Watson, 2005)。该原则认为，在有限制的情景中，行为表现者通过可得到的最合理的方法实现目标状态，这也正是意向性动作生效的方式。合理动作原则包含三个成分：动作、目标状态和情景限制。Király, Jovanovic, Prinz, Aschersleben 和 Gergely (2003)认为合理动作原则包含两个前提假定。第一个假定认为，动作的基本功能是带来客观环境的特定变化，这表明动作结果应该包括环境状态的明显变化，在此重点强调了三个成分中的“目标状态”。第二个假定则认为，在情景限制下主体会利用其可以获得的最有效方法。该假定强调情景限制改变时，主体为了高效地实现目标会采用不同的动作。这一假定更加重视三个成分中的“动作”和“情景限制”。也就是说，使用合理动作原则帮助人们进行意图识别和检测是基于这样的推论：如果动作可以体现主体意图，那么，对主体在限制情景下动作及其带来的现实状态改变的察觉，可以帮助人们确认动作的隐含意图。已有研究证实动作合理性在意图识别上的确是有效的参考指标(Bonchek-Dokow & Kaminka, 2014)。

除了对影响动作意图测量的各个因素的研究之外，也有研究者对动作意图识别过程中各个子过程的检测进行区分。研究者认为意图识别有两个核心过程，分别为意图检测和意图预测。这两个阶段由于其具体目的不同，研究侧重点也有所不同。动作意图检测是为了确定意图的存在，这一阶段主要分析观察到的动作序列的每一点；而动作意图预测则是为了确定意图内容，在这一阶段需要在时间进程上向前推进，从观察到的行为导致的最终状态出发，推论行为者的预期目标(Bonchek-Dokow & Kaminka, 2014)。

4.2 不同水平信息对动作意图的影响

通过中介因素研究动作意图识别时，对行为的理解往往同时涉及低级层次和高级层次两个水平。其中低级层次主要包括人体检测与跟踪、动作识别、手势识别等，而高级层次主要考虑背景因素的影响等(徐光祐, 曹媛媛, 2009)。同时，还需要考虑背景–刺激物关系及刺激物–刺激物关系对于动作意图识别的影响。背景–刺激物关系常常通过自上而下的知识经验影响动作意图识别；换言之，人们首先需要有关于动作物理特征(包括动作物理特征和动作的序列性信息)和语义特征(包括场景与动作的匹配性问题)的知识经验，随后根据知识经验及所观察到的信息来推论或识别动作的隐含意图。另外，刺激物–刺激物关系对动作意图识别的影响也受到自上而下的知识经验的影响。刺激物与刺激物之间联系的构建一般都是与人们已有的知识系统息息相关的。但是，无论是背景–刺激物关系所依赖的物理特征，还是刺激物–刺激物关系中刺激物的各种物理特征，都是直接通过自下而上的识别过程得到的。这与计算机识别的方式是一致的。

4.3 动作意图识别的计算机模型

近年来，如何利用背景信息促进视觉识别不仅引起了场景知觉领域研究者的关注，而且也成为机器视觉(computer vision)研究的一个重要内容。研究者发现，背景信息可以用于动作分类(Marszalek, Laptev, & Schmid, 2009)、场景及其包含的刺激物的识别(Divvala, Hoiem, Hays, & Efros, 2009; Rabinovich, Vedaldi, Galleguillos, Wiewiora, & Belongie, 2007)等。然而，对于复杂场景中的动

作姿势，仍然没有非常有效的方法予以识别。那么，能不能基于合理动作原则，以及场景不同水平的信息，对场景中的动作意图进行训练学习并计算模拟呢？

Yao 和 Fei-Fei (2010)在前人研究的基础上，提出了动作意图的计算机识别模型。该模型试图在刺激物觉察和动作姿势估计之间建立联系，并假设共同背景(mutual contexts)对于二者之间关系的理解具有重要的影响作用，并且可以促进动作姿势的估计以及刺激物的觉察。在真实的生活场景中，每一个具体的人−物互动(human-object interaction, HOI)活动都是具有特殊性的，都是不同于其它任何活动的，因此, Yao 和 Fei-Fei 所建立的实际上是一个将 HOI 活动场景分解为活动类别、刺激物对象和身体姿势的分层随机场(hierarchical random field)模型。其中身体姿势又可以分解成身体的不同部位，而每一个身体部位和刺激物对象则可以表示为相应的视觉特征，其他潜在变量则可以通过机器训练学习获得。

Yao 和 Fei-Fei (2012)在之后的研究中发现，该模型可以用于检测图像中人的身体姿势以及与其有交互作用的对象，并且发现利用人体动作姿势更有利于促进相关刺激物的识别，其检测性能显著优于词汇袋的方法，也略优于 Gupta, Kembhavi 和 Davis (2009)提出的基于背景场景环境进行检测的方法。更重要的是，该模型的应用将“刺激物是什么”的识别转向“刺激物是用来做什么”的识别(Koppula, Gupta, & Saxena, 2013)，这对于真实场景中刺激物识别的研究而言具有重要的引导作用。

5 小结与展望

20 世纪 30 年代以来，场景及场景中刺激物的识别始终是研究者关注的核心理论问题之一。然而，与自然场景不同，社会场景中人的注视方向、动作行为等都影响着观察者的信息加工、行为决策等(Kingstone, Smilek, Ristic, Friesen, & Eastwood, 2003; Gibson & Kingstone, 2006)；而动作意图的识别与检测也已成为社会场景知觉及其语义获得的主要研究内容。在未来的相关研究中，个体动作意图识别能力的差异性及其发展、真实场景中动作意图识别的文化差异、机器视觉研究的优化以及计算机模型的修正等可能是该领域未来研究的重要方向。

5.1 个体动作意图识别能力的差异性及其发展

有研究者认为动作意图识别是一种认知他人计划、目的的能力(Sukthankar et al., 2014)。从这个角度而言，动作意图识别作为一种个体能力，不同个体由于其生活环境、知识经验等的差异，识别他人动作意图的能力也可能存在着不同。因此，对动作意图识别能力个体差异的研究，可能是未来研究中十分重要的方向。基于此，从能力发展的角度考虑这种差异，与年龄相关的信息加工能力差异是否对动作意图识别能力的差异有所贡献？动作意图识别能力的发展是阶段性的或者连续性的，相关问题的探讨有助于人们在儿童发展的适当阶段，通过适当的引导教育，促进他们动作意图识别能力的发展。

5.2 真实场景中动作意图识别的文化差异

场景作为一种真实环境信息，可以作为信息载体，为人们提供各种信息；同时，场景提供的信息对不同文化背景下的个体又具有不同的心理含义。如，西方饮食文化与中国饮食文化背景下的个体对于“使用筷子为他人夹菜”这一行为会得到不同的意图推论。因此，对于真实场景中人物的动作意图的识别，除了个体差异之外，也可能存在着深刻的文化差异。在全球化的背景下，不同文化之间的交流愈加频繁，基于场景中人们的动作考察意图识别的文化差异具有重要的现实意义。诸如在怎样的场景下动作意图识别具有人类的普遍性，而在怎样的场景下动作意图又具有明显的文化差异；动作意图识别过程中的文化差异是由于个体人的原因，还是文化环境的原因等问题，仍然需要研究者进一步探讨。

5.3 机器视觉研究的优化以及计算机模型的修正

在机器视觉的研究中，动作意图识别也是一个应用广泛的课题。例如，在智能监控领域使用智能化的视频监控手段并使用计算机帮助人类进行分析和监控，可以有效避免人工监控中存在的效率低、耗费大，以及可能有遗漏的安全隐患的问题(杜有田，陈峰，徐文立，李永彬, 2007)。此外，基于机器视觉的人体运动分析研究可以通过提取运动员关节位置、角度、速度等信息，并通过对这些数据信息的分析和处理，指导下一步的训练(黎洪松，李达，2009)。值得注意的是，机器视觉

下的动作意图研究本就是建立在对人的意图的研究基础之上的，因此，从认知神经科学的角度对动作意图识别的深入研究(Wang et al., 2010; Wang, Zheng, Lin, Wu, & Shen, 2011)，以及对动作意图识别过程中的内在心理机制的探索(Meltzoff, 2007)，对于推进机器视觉动作意图识别能力的优化，以及计算机模型的修正都是非常必要的。

除此之外，场景背景对动作意图识别并不总是发挥正性作用。如果是嘈杂或混乱的情景，可能对隐含意图的识别产生负面影响(Klaser, Marszek, Laptev, & Schmid, 2010)；而且，一个场景可能包括不同的动作，如果不能提供有用的信息来区分这些动作，对于识别具有隐含意图的动作行为也有负面影响(Ziaeefard & Bergevin, 2015)。因此，从场景背景可能产生影响的性质角度出发，探索真实情景中动作意图识别也是有必要的。

参考文献

白学军，康廷虎，闫国利. (2008). 真实情景中刺激物识别的理论模型与研究回顾. *心理科学进展, 16*(5), 679–686.

陈亚萍，李晓东. (2013). 婴儿动作理解的研究回顾与展望. *心理科学进展, 21*(4), 671–678.

杜有田，陈峰，徐文立，李永彬. (2007). 基于视觉的人的运动识别综述. *电子学报, 35*(1), 84–90.

康廷虎，白学军. (2008). 真实情景知觉中注视控制的研究进展. *西北师大学报(社会科学版), 45*(4), 107–111.

李红，何磊. (2003). 儿童早期的动作发展对认知发展的作用. *心理科学进展, 11*(3), 315–320.

黎洪松，李达. (2009). 人体运动分析研究的若干新进展. *模式识别与人工智能, 22*(1), 70–78.

王福兴，田宏杰，申继亮. (2009). 场景知觉及其研究范式. *心理科学进展, 17*(2), 268–277.

徐光祐，曹媛媛. (2009). 动作识别与行为理解综述. *中国图象图形学报, 14*(2), 189–195.

Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences, 5*(4), 171–178.

Bonchek-Dokow, E., & Kaminka, G. A. (2014). Towards computational models of intention detection and intention prediction. *Cognitive Systems Research, 28*, 44–79.

Cacippo, J. T., Berntson, G. G., & Decety, J. (2010). Social neuroscience and its relationship to social psychology. *Social Cognition, 28*(6), 675–685.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior & Development, 21*(2), 315–330.

Catmur, C. (2015). Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Consciousness and Cognition, 36*, 426–433.

Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2007). Predicting human gaze using low-level saliency combined with face detection. In *Proceedings of the 20th International conference on neural information processing systems* (pp. 241–248). Vancouver, British Columbia, Canada: Curran Associates Inc..

Choi, D. (2013). Design and implementation of context awareness system for abnormal behavior detection. Unpublished results. Dept of Computer Science, Gachon University.

Delaitre, V., Sivic, J., & Laptev, I. (2011). Learning person-object interactions for action recognition in still images. In *Proceedings of the 24th international conference on neural information processing systems* (pp. 1503–1511). Granada, Spain: Curran Associates Inc..

den Ouden, H. E. M., Frith, U., Frith, C., & Blakemore, S. J. (2005). Thinking about intentions. *NeuroImage, 28*(4), 787–796.

Desai, C., Ramanan, D., & Fowlkes, C. (2010). Discriminative models for static human-object interactions. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 9–16.

Divvala, S. K., Hoiem, D., Hays, J. H., & Efros, A. A. (2009). An empirical study of context in object detection. In *IEEE conference on computer vision and pattern recognition* (pp. 1271–1278). Miami, Florida, USA: IEEE.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*(3), 316–355.

Gibson, B. S., & Kingston, A. (2006). Visual attention and the semantics of space. *Psychological Science, 17*(7), 622–627.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology*. Hoboken, NJ: John Wiley & Sons Inc.

Gupta, A., Kembhavi, A., & Davis, L. S. (2009). Observing human-object interactions: Using spatial and functional compatibility for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(10), 1775–1789.

Henderson, J. M. (2005). Introduction to real-world scene perception. *Visual Cognition, 12*(6), 849–851.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*(1), 243–271.

Jang, Y. M., Lee, S., Mallipeddi, R., Kwak, H. W, & Lee, M. (2013). Intent probing monitoring system based on eye movement analysis and probability. *The Institute of Electronics and Information Engineers, 36*(1), 1518–1521.

Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science, 12*(5), 176–180.

Király, I., Jovanovic, B., Prinz, W., Aschersleben, G., & Gergely, G. (2003). The early origins of goal attribution in infancy. *Consciousness and Cognition, 12*(4), 752–769.

Klaser, A., Marszek, M., Laptev, I., & Schmid, C. (2010). Will person detection help bag-of-features action recognition? *European Journal of Neuroscience, 23*(2), 365–373.

Koppula, H. S., Gupta, R., & Saxena, A. (2013). Learning human activities and object affordances from RGB-D videos. *The International Journal of Robotics Research, 32*(8), 951–970.

Marszalek, M., Laptev, I., & Schmid, C. (2009). Actions in context. In *IEEE conference on computer vision and pattern recognition* (pp. 2929–2936). Miami, FL: IEEE.

Meltzoff, A. N. (2007). The “like me” framework for recognizing and becoming an intentional agent. *Acta Psychologica, 124*(1), 26–43.

Muehlhaus, J., Heim, S., Altenbach, F., Chatterjee, A., Habel, U., & Sass, K. (2014). Deeper insights into semantic relations: An fMRI study of part-whole and functional associations. *Brain & Language, 129*, 30–42.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11*(12), 520–527.

Park, H., Lee, S., Lee, M., Chang, M. S., & Kwak, H. W. (2016). Using eye movement data to infer human behavioral intentions. *Computers in Human Behavior, 63*, 796–804.

Rabinovich, A., Vedaldi, A., Galleguillos, C., Wiewiora, E., & Belongie, S. (2007). Objects in context. In *IEEE 11th international conference on computer vision* (pp. 1–8). Rio de Janeiro, Brazil: IEEE.

Ruys, K. I., & Aarts, H. (2010). When competition merges people’s behavior: Interdependency activates shared action representations. *Journal of Experimental Social Psychology, 46*(6), 1130–1133.

Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: The role of movement information. *Cognition, 119*(2), 242–252.

Satpute, A. B., Fenker, D. B., Waldmann, M. R., Tabibnia, G., Holyoak, K. J., & Lieberman, M. D. (2005). An fMRI study of causal judgments. *European Journal of Neuroscience, 22*(5), 1233–1238.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences, 10*(2), 70–76.

Sukthankar, G., Geib, C., Bui, H. H., Pynadath, D., & Goldman, R. P. (2014). *Plan, activity, and intent recognition: Theory and practice* (pp. 19–20). London: Morgan Kaufmann.

Wang, Y. W., Lin, C. D., Yuan, B., Huang, L., Zhang W. X., & Shen, D. L. (2010). Person perception precedes theory of mind: An event related potential analysis. *Neuroscience, 170*(1), 238–246.

Wang, Y. W., Zheng, Y. W., Lin, C. D., Wu, J., & Shen, D. L. (2011). Electrophysiological correlates of reading the single- and interactive-mind. *Frontiers in Human Neuroscience, 5*. doi: 10.3389/fnhum.2011.00064

Watson, J. S. (2005). The elementary nature of purposive behavior: Evolving minimal neural structures that display intrinsic intentionality. *Evolutionary Psychology, 3*(1), 24–48.

Yao, B. P., & Fei-Fei, L. (2010). Modeling mutual context of object and human pose in human-object interaction activities. In *Proceedings of the 23th IEEE conference on computer vision and pattern recognition* (pp. 17–24). San Francisco, CA, USA: IEEE.

Yao, B. P., & Fei-Fei, L. (2012). Recognizing human-object interactions in still images by modeling the mutual context of objects and human poses. *IEEE Transactions on Pattern Analysis and machine Intelligence, 34*(9), 1691–1703.

Ziaeefard, M., & Bergevin, R. (2015). Semantic human activity recognition: A literature review. *Pattern Recognition, 48*(8), 2329–2345.

Recognition of action and intention in real-world scene perception

KANG Tinghu; XUE Xi

(*Visual Cognition Lab, School of Psychology, Northwest Normal University, Lanzhou* 730070, *China*)

Abstract: A social scene plays a crucial part in the real physical world that people live in. In social scene perception studies, recognition of actions and associated intentions can be influenced not only by the background information of the scene, but can also be related to the object of an action. Therefore, researchers could follow the relationships between the background and an object, or among various objects for analyzing the mechanism of action recognition. However, to detect and recognize an action and its associated intention, researchers could also employ semantics restriction and physical baffle of scene, and incorporate the principle of rational action for studying the biological signs following an action. In the field of machine vision, new research is emerging on models of computer recognition that are based on human-object interaction. In the future, researchers can consider the development of action and intention identification capacity, and can study the differences among individuals of various cultures for improving the studies conducted in this field of research.

Key words: social scene; action and intention; scene perception; model of computer recognition.